



On the link between credibility and frequency premium

Jean Pinquet, Guillén Montserrat, Catalina Bolancé

► To cite this version:

Jean Pinquet, Guillén Montserrat, Catalina Bolancé. On the link between credibility and frequency premium. 2007. hal-00243063

HAL Id: hal-00243063

<https://hal.science/hal-00243063>

Preprint submitted on 6 Feb 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



ÉCOLE POLYTECHNIQUE

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE



On the link between credibility and frequency premium

Jean Pinquet
Montserrat Guillén
Catalina Bolancé

September 2007

Cahier n° 2007-17

LABORATOIRE D'ECONOMETRIE

1 rue Descartes F-75005 Paris

(33) 1 55558215

<http://ceco.polytechnique.fr/>

<mailto:lyza.racon@shs.poly.polytechnique.fr>

On the link between credibility and frequency premium¹

Jean Pinquet²
Montserrat Guillén³
Catalina Bolancé⁴

September 2007

Cahier n° 2007-17

Résumé: Ce papier remet en question l'hypothèse d'équidistribution des effets aléatoires dans un modèle de risque fréquence. Deux modèles sont présentés, qui utilisent des liens paramétriques et non paramétriques entre la variance de l'effet aléatoire et le risque fréquence. Ils sont estimés sur un portefeuille de contrats espagnols d'assurance automobile, pour lesquels le lien précité est décroissant. Des conclusions en sont tirées pour la crédibilité et les coefficients bonus-malus.

Abstract: This paper questions the equidistribution assumption for the random effects in a frequency risk model. Two models are presented, which use parametric and nonparametric links between the variance of the random effect and frequency risk. They are estimated on a Spanish automobile insurance portfolio, for which a decreasing link is obtained. Conclusions are drawn for credibility and bonus-malus coefficients.

Mots clés : Estimateurs à noyaux. Processus gamma.

Key Words : Kernel-based estimators, gamma processes.

Classification JEL: C14, C25.

¹ The authors acknowledge financial support from FEDER and from the Spanish Ministry of Education and Science.

² Université Paris 10 & Laboratoire d'Econométrie, Ecole Polytechnique. E-mail address: jean.pinquet@polytechnique.edu

³ Dept. d'Econometria, Estadística i Economia Espanyola, Universitat de Barcelona. E-mail address: mguillen@ub.edu

⁴ Dept. d'Econometria, Estadística i Economia Espanyola, Universitat de Barcelona. E-mail address: bolance@ub.edu

1 Introduction

Usual actuarial models assume that random effects used in distributions of the number of claims are identically distributed. In this framework, the credibility granted to the history of the policyholder increases with frequency premium. Credibility is the no-claim discount for a claimless history, and the increasing relationship between an actuarial bonus and an estimated risk level is quoted in various papers (e.g. Dionne and Vanasse, 1989).

There is however empirical evidence of a decreasing link between the variance of the random effect and frequency risk for automobile insurance data (see Sections 3 and 4 with results obtained from a Spanish portfolio). In other words, residual relative heterogeneity on claims number distributions is more important for low risks.

In this paper, the variance of the random effect applied to Poisson distributions is conditioned on frequency risk and hence on the rating factors. First, we retain a local estimation approach of a nonparametric link between the variance and the frequency.¹ Section 2 summarizes the main properties of kernel-based estimators in generalized linear models. This approach is used in Section 3 to estimate the nonparametric link. Second, a parametric power link is specified and estimated in Section 4 from the negative binomial model. Consequences on credibility derivation are drawn in Section 5. Section 6 concludes and a mathematical appendix contains some mathematical details.

The main empirical finding is that the link between credibility (or no-claim discount) and frequency premium is lower when the equidistribution assumption of the random effect is relaxed than in the usual model. An opposite result is obtained for the increase in premium after a claim.

¹Local estimation techniques can also be used for prediction on time series (see Qian, 2000, for applications to insurance).

2 Kernel estimators in generalized linear models:

The index model

Generalized linear models for a response variable Y and a vector of regressors X ($X \in \mathbb{R}^k$) assume that

$$E(Y | X = x) = f(x'\beta), \beta \in \mathbb{R}^k. \quad (1)$$

The function f is a link between an index $x'\beta$ (i.e. a scalar product between regressors and parameters) and the expectation of the response variable. In the literature on generalized linear models, the link usually refers to the reciprocal of f , but we retain here the function which is estimated in the first place. For identifiability purposes detailed afterwards, we suppose that the intercept is not included as a regressor in (1) and that the distribution of X is nondegenerate in \mathbb{R}^k . For basic generalized linear models, f is given and then an intercept must be included in the regression. For a count data model, f is usually the exponential function. If f is unknown in the specification so that an estimation is required, equation (1) is referred to as an index model (see Härdle et al., 1997). In that case, there is an obvious identifiability conflict between β and f in equation (1). Only the line $\mathbb{R}\beta$ can be identified from the data. In other words, what is identified is the conditional expectation, assumed to be constant on affine hyperplanes of \mathbb{R}^k which are orthogonal to a given vector. For a given value of β , a nonparametric estimation of $f(s)$ can be based on local weighted averages of the response variable, with weights which decrease with the distance between s and the individual values of the index.

A first estimation of index models can be obtained from a parametric specification of the distribution of Y defined conditionally on X . Let us assume that we have in that case

$$E(Y | X = x) = f_0(c + x'b); \quad c \in \mathbb{R}, \quad b \in \mathbb{R}^k, \quad (2)$$

with f_0 a given link function. Let \hat{b} be the maximum likelihood estimator of b . The conditional expectation defined in (1) can be estimated with a kernel estimator.

In a sampling model on (X, Y) with n observations $(x_i, y_i)_{i=1, \dots, n}$, an estimator of $E(Y | X)$ of the Nadaraya-Watson type is obtained from a kernel K (usually an even probability density function) and a bandwidth h in the following way:

$$\hat{E}(Y | X) = \frac{\sum_{i=1}^n y_i K_h \left(X' \hat{b} - x_i' \hat{b} \right)}{\sum_{i=1}^n K_h \left(X' \hat{b} - x_i' \hat{b} \right)}, \quad K_h(u) = \frac{K\left(\frac{u}{h}\right)}{h}. \quad (3)$$

The bandwidth is a smoothing parameter. The closer it is to zero, the more estimation is performed on a local basis. The estimation given in (3) exhibits an invariance property as it only depends on \hat{b}/h .

A suitable value of h can be derived from a cross-validation method similar to that proposed in Härdle (1990) for the Nadaraya-Watson kernel estimator. It is equal to

$$\arg \min_h CV(\hat{b}, h) = \sum_{i=1}^n (y_i - \hat{E}_{-i}(Y_i))^2, \quad (4)$$

where $\hat{E}_{-i}(Y_i)$ is the leave-one-out estimator² defined by

$$\hat{E}_{-i}(Y_i) = \frac{\sum_{j \neq i} y_j K_h \left(x_i' \hat{b} - x_j' \hat{b} \right)}{\sum_{j \neq i} K_h \left(x_i' \hat{b} - x_j' \hat{b} \right)} = \frac{\sum_{j \neq i} y_j K \left((x_i - x_j)' \frac{\hat{b}}{h} \right)}{\sum_{j \neq i} K \left((x_i - x_j)' \frac{\hat{b}}{h} \right)}. \quad (5)$$

This estimation of the conditional expectation $E(Y | X)$ is not necessarily consistent since it is derived from a wrong link function (f_0 instead of f). Two results are worth mentioning on this issue.

- On one hand, a consistent estimator of the conditional expectation can be obtained from the cross-validation criterion defined in (4) and (5). Replacing

²The individual for which the non parametric expectation is derived must be withdrawn from the computation, otherwise the bandwidth would converge towards zero in the minimization of the cross-validation criterion.

\hat{b} by b in (4) defines a function $CV(b, h)$ which can be minimized with respect to b and h . The optimal values of b and h are plugged into the expression of the estimated conditional expectation given in (3). Sufficient conditions for the consistency of the estimation are given in Härdle et al. (1993). However the minimization of the cross-validation criterion is cumbersome, since it necessitates a double sum on individuals. Besides, equation (5) shows that only b/h is identified. Hence an identifying constraint needs to be added in the estimation.

- On the other hand, the maximum likelihood estimation of a parametric model given in the first place can lead to consistent estimation of the conditional expectation under conditions which are first related to the distribution of the regressor X (see Li and Duan, 1989, and the appendix). Owing to the identification issue mentioned before, consistency means that \hat{b} converges towards a limit b_0 which belongs to the line $\mathbb{R}\beta$.

3 Kernel estimators for the variance of the random effect in a Poisson model

Let us consider cross-section data. The policyholders in the portfolio are indexed by $i = 1, \dots, p$. All the risk exposure durations are supposed equal (they are equal to one year in our empirical study). Frequency risk must be expressed for a time unit, otherwise the results on the link investigated in this paper would not be coherent with respect to period aggregation. We denote n_i as the number of claims reported by policyholder i and x_i as the vector of regression components. If U_i is the random effect, the distribution of N_i in the Poisson model with random effects is obtained

from the expectation taken with respect to the random effect U_i

$$P[N_i = n_i] = E[P_{\lambda_i U_i}(n_i)]; \lambda_i = \exp(c + x_i' b); P_{\lambda}(n) = \exp(-\lambda) \frac{\lambda^n}{n!}.$$

If the equidistribution assumption of the random effects is relaxed, we can link their variance and frequency risk and write for instance

$$E(U_i) = 1; V(U_i) = \sigma^2(\lambda_i), \lambda_i = E(N_i). \quad (6)$$

The function σ^2 must have nonnegative values. We will investigate a power link in Section 4 but we first let the data speak from a nonparametric estimation of σ^2 . The starting point is the usual moment-based estimator

$$V(U_i) = \frac{V(N_i) - E(N_i)}{E^2(N_i)} = \frac{E(N_i^2) - E^2(N_i) - E(N_i)}{E^2(N_i)}. \quad (7)$$

The nonparametric link is then obtained with an index model strategy described in Section 2. First, an estimation is performed from maximum likelihood estimation of the Poisson model with regression components

$$N_i \sim P(\lambda_i), \lambda_i = \exp(c + x_i' b).$$

For each policyholder i , we obtain the index $s_i = \widehat{c} + x_i' \widehat{b}$ and the parametric frequency premium $\widehat{E}^0(N_i) = \exp(s_i)$. Then the variance of the random effect is estimated from equation (7) and from nonparametric estimators of $E(N_i^m)$, $m = 1, 2$. We use equation (3), with $Y = N^m$.

In what follows, we retain a Gaussian kernel. Hence K_h is the density of a $N(0, h^2)$ distribution. The bandwidth h_m retained for the estimation of $E(N_i^m)$ ($m = 1, 2$) is obtained with the leave-one-out approach (see equations (4) and (5)). From equations (7) and (3), a nonparametric estimator of $V(U_i)$ is

$$\widehat{V}(U_i) = \frac{\widehat{E}(N_i^2) - \widehat{E}(N_i) - \widehat{E}^2(N_i)}{\widehat{E}^2(N_i)}.$$

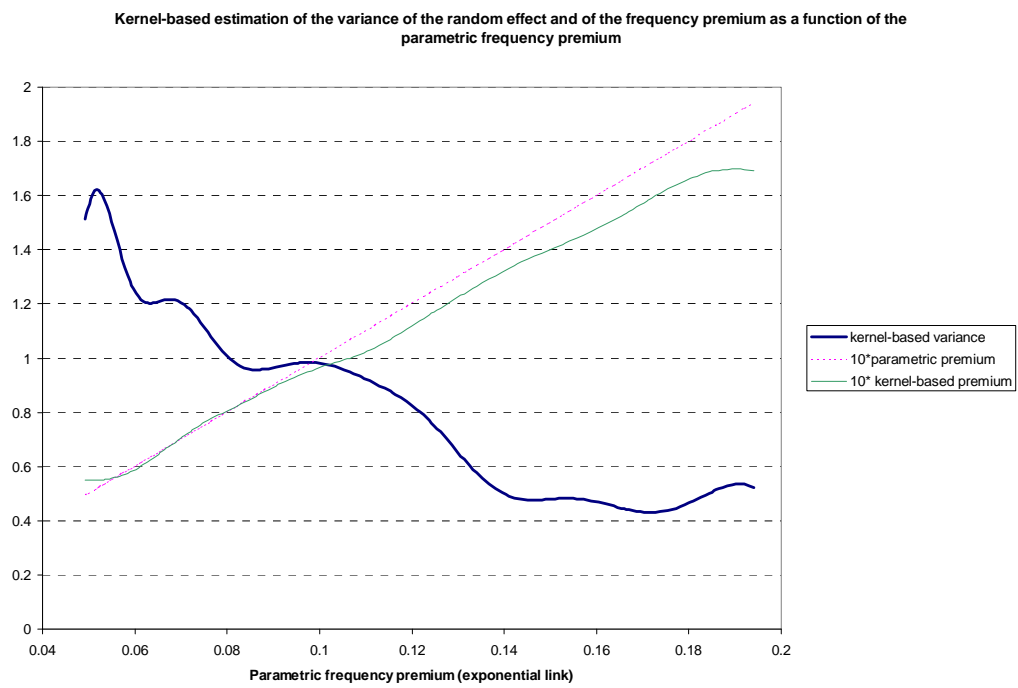


Figure 1:

The estimated variance of the random effect is then a nonparametric function of the frequency premium $\hat{E}^0(N_i)$.

Figure 1 presents the empirical link between $\hat{V}(U)$, $\hat{E}(N)$ and $\hat{E}^0(N)$ for a Spanish portfolio containing 80,994 policyholders observed during one year.³ The response variable is the number of claims at fault. The rating factors are the gender, the geographical area, the age of the driving licence, the age of the policyholder and her seniority as a customer of the insurance company, the coverage level and the power of the vehicle. Figure 1 is obtained with the two step estimation approach described after equation (7). A direct estimation of the regression parameters jointly with the bandwidth from the cross-validation criterion led to very similar results. For instance, the cosine between the two estimations linked to the regressors is equal to 0.9996, which indicates almost perfect colinearity. As only b/h is identified (see equation (5) and the following comments), this means that the estimations are almost equivalent. This result was not obvious *ex ante*, as consistency with link misspecification holds under assumptions on the distribution of regressors which are not necessarily fulfilled with the qualitative variables used in our regressions (see the appendix).

Figure 1 exhibits a decreasing link between the local estimation for the variance of the random effect and the frequency premium. This result was confirmed on other data bases. A power link between the variance of the random effect and frequency risk is estimated in the following section.

³The bandwidths retained for the first and second order moment of the response variable from the cross-validation criterion given in (4) are equal to $h_1 = h_2 = 0.009$. Notice that the two estimations of the frequency are close to each other around the average value (equal to 0.07), whereas the kernel-based estimation is lower than the fully parametric estimation for large values of the frequency.

4 Negative binomial model with a power link between the variance of the random effect and frequency risk

In this section, we use a modified version of the negative binomial model with regression components, in which we include a power link between the variance of the random effect and frequency risk. Hence we have a parametric model which can reflect the decreasing link observed in Figure 1. This model was proposed by Winkelmann and Zimmermann (1991).

We keep the notations of Section 3, and we suppose that the random effect follows a gamma distribution. We write

$$U_i \sim \gamma(a_i, a_i); \quad a_i = a\lambda_i^{-e}, \quad \lambda_i = E(N_i) = \exp(c + x_i' b).$$

The gamma distribution is indexed by a shape and a scale parameter. We have

$$E(U_i) = 1; \quad V(U_i) = \frac{1}{a_i} = \frac{\lambda_i^e}{a}, \quad (8)$$

and the parameter e is the elasticity of $V(U_i)$ with respect to frequency risk. The equidistribution assumption for the random effects means that this elasticity is null.

The variance of the random effect specified by (8) is an exponential function of the index $x_i' b$. Such a link was investigated for a linear model by Harvey (1976).

The likelihood is the usual one for the negative binomial model, hence

$$P[N_i = n_i] = E[P_{\lambda_i U_i}(n_i)] = \frac{a_i^{a_i} \lambda_i^{n_i}}{(\lambda_i + a_i)^{n_i + a_i}} \frac{\Gamma(n_i + a_i)}{\Gamma(a_i) \Gamma(n_i + 1)}.$$

The maximum likelihood estimators for the Spanish portfolio are the following.

$$\hat{a} = 8.05; \quad \hat{e} = -0.839; \quad \frac{\hat{e}}{\widehat{\sigma_{\hat{e}}}} = -2.141. \quad (9)$$

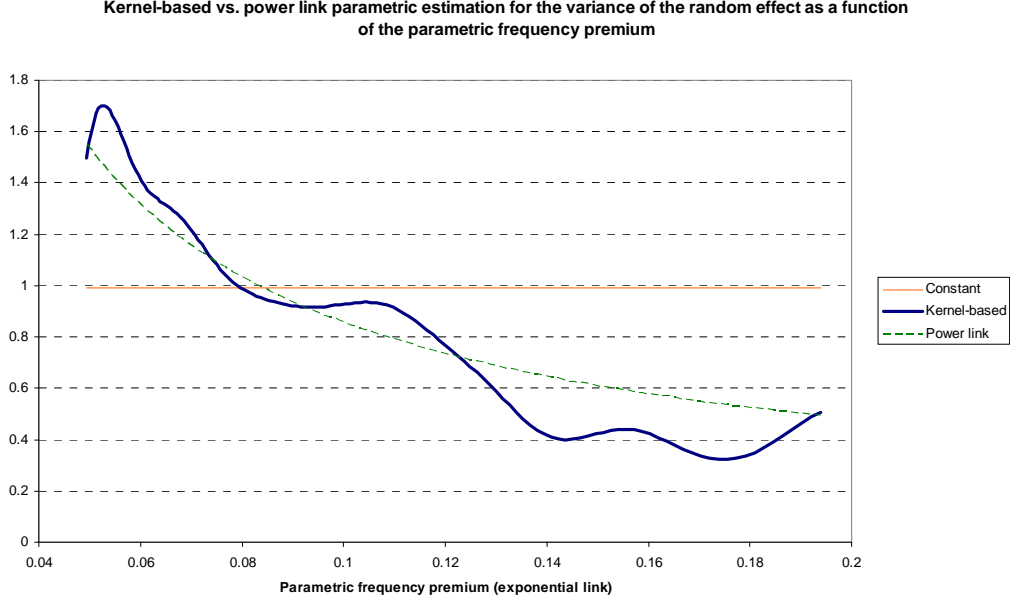


Figure 2:

The usual semiparametric estimator for a constant variance of the random effect is equal to

$$\hat{\sigma}^2 = \hat{V}(U) = \frac{\sum_i (n_i - \hat{\lambda}_i)^2 - n_i}{\sum_i \hat{\lambda}_i^2} = 0.985.$$

Figure 2 plots two estimated variances of the random effects, defined as a power or a kernel-based function of the parametric frequency premium, together with the constant estimation.

5 Applications to credibility predictors

Linking the variance of the random effect and frequency risk raises a difficulty in a longitudinal data analysis. Since frequency risk varies with time, the random effect

cannot be supposed time-independent in the prediction. A solution is to derive the random effects from a stochastic process defined in continuous time, and to link the time index of the process with frequency risk. Let us detail an example from a gamma process.

If data are longitudinal, we simply replace i by the pair i, t (where t indexes the periods) in the preceding equations. For instance, the random effects are defined as follows

$$U_{i,t} \sim \gamma(a_{i,t}, a_{i,t}) \Leftrightarrow a_{i,t} U_{i,t} \sim \gamma(a_{i,t}). \quad (10)$$

We obtain the random effects $U_{i,t}$ from a gamma process $(G_{i,a})_{a \geq 0}$, which is defined by three properties:

1) $G_{i,0} \equiv 0$; 2) The process has independent increments; 3) The increments follow gamma distributions, with a parameter equal to the difference in dates (i.e. $G_{i,a_2} - G_{i,a_1} \sim \gamma(a_2 - a_1) \forall a_1, a_2, a_2 > a_1 \geq 0$). For instance, we have $G_{i,a} \sim \gamma(a)$.

We have a particular type of Levy process with indefinitely divisible distributions. This process exists from the well known property on gamma distributions

$$\gamma(a_1) * \gamma(a_2) = \gamma(a_1 + a_2) \quad (a_1, a_2 \geq 0)$$

and from Kolmogorov's theorem. The gamma process can also be seen as a limit of compound Poisson processes (see Dufresne et al., 1991, for definitions and applications to ruin theory).

Suppose that we have the link $V(U_{i,t}) = \sigma^2(\lambda_{i,t})$ given in equation (6) between the variance of the random effect and the frequency risk. For the distributions given in equation (10), we have $V(U_{i,t}) = 1/a_{i,t}$. This leads us to define the random effects as follows

$$U_{i,t} = \sigma^2(\lambda_{i,t}) \times G_{i,1/\sigma^2(\lambda_{i,t})}. \quad (11)$$

Then we have

$$\lambda_{i,t_1} = \lambda_{i,t_2} \Rightarrow U_{i,t_1} = U_{i,t_2}.$$

We can consider for instance the power link $\sigma^2(\lambda_{i,t}) = a \times \lambda_{i,t}^{-e}$ retained in Section 4.

Let us predict frequency risks with a linear credibility approach (Bühlmann, 1967). The bonus-malus coefficient for the second period of the policyholder i is derived from an affine probabilistic regression of U_{i2} with respect to N_{i1} . We have

$$\widehat{u}_{i2} = 1 + \frac{\widehat{Cov}(U_{i2}, N_{i1})}{\widehat{V}(N_{i1})}(n_{i1} - \widehat{\lambda}_{i1}),$$

with

$$\widehat{Cov}(U_{i2}, N_{i1}) = \widehat{\lambda}_{i1} \widehat{Cov}(U_{i2}, U_{i1}); \quad \widehat{V}(N_{i1}) = \widehat{\lambda}_{i1} + \widehat{\lambda}_{i1}^2 \widehat{V}(U_{i1}).$$

Since the gamma process $G_{i,t}$ has independent increments, we have that

$$Cov(G_{i,a_1}, G_{i,a_2}) = V(G_{i,\min(a_1,a_2)}) = \min(a_1, a_2).$$

From (11) and the last equation, we obtain

$$\widehat{Cov}(U_{i2}, U_{i1}) = \frac{\widehat{\sigma}^2(\widehat{\lambda}_{i1}) \times \widehat{\sigma}^2(\widehat{\lambda}_{i2})}{\max(\widehat{\sigma}^2(\widehat{\lambda}_{i1}), \widehat{\sigma}^2(\widehat{\lambda}_{i2}))}.$$

Then the linear credibility predictor is the weighted average

$$\widehat{u}_{i2} = 1 + cred_i \left(\frac{n_{i1}}{\widehat{\lambda}_{i1}} - 1 \right) = (1 - cred_i) + \left(cred_i \times \frac{n_{i1}}{\widehat{\lambda}_{i1}} \right).$$

The credibility $cred_i$ granted to the first period is equal to

$$cred_i = \frac{\widehat{\lambda}_{i1} \times \widehat{Cov}(U_{i2}, N_{i1})}{\widehat{V}(N_{i1})} = \frac{\widehat{\lambda}_{i1} \times \widehat{\sigma}^2(\widehat{\lambda}_{i1})}{1 + (\widehat{\lambda}_{i1} \times \widehat{\sigma}^2(\widehat{\lambda}_{i1}))} \times \frac{\widehat{\sigma}^2(\widehat{\lambda}_{i2})}{\max(\widehat{\sigma}^2(\widehat{\lambda}_{i1}), \widehat{\sigma}^2(\widehat{\lambda}_{i2}))}. \quad (12)$$

Let us comment equation (12). The first component of the product which defines the credibility is the usual formula, with a variance of the random effect $\widehat{\sigma}^2$ which depends on the frequency premium. As the function $\widehat{\sigma}^2$ decreases with the frequency premium on our data, this component of the credibility increases less than with the usual formula. It might even decrease if the estimated elasticity between the variance of the random effect and frequency risk (defined globally with a power link function

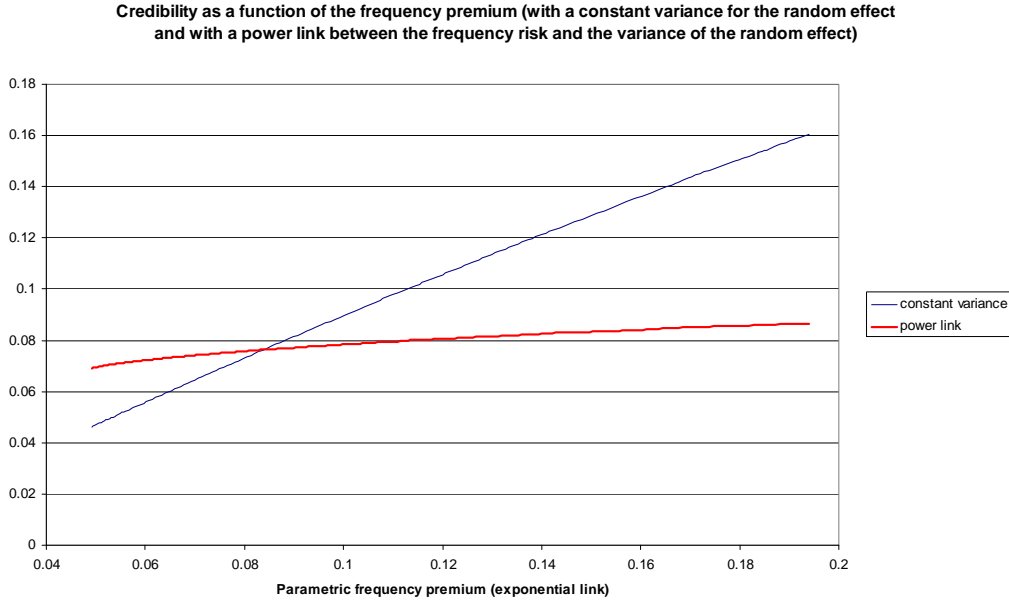


Figure 3:

or locally with a nonparametric link) was less than -1. The second component of the credibility in equation (12) is less than one. Frequency premiums do not vary much for a given policyholder from one period to the following⁴, and this component remains close to one. All the policyholders of our sample stay in the portfolio at the second period.⁵ On our data, the average of the second component of (12) derived from the power link estimated in (9) is equal to 0.99.

In Figure 3, we plot two derivations of the credibility as a function of the frequency premium $\widehat{\lambda}_{i1}$. First, credibility is derived from the usual formula with a

⁴A thorough analysis of stochastic migration between risk levels during different periods is given in Brouhns et al. (2003).

⁵They are actually present during seven years. See Bolancé et al. (2003) for an investigation of the whole panel data set.

constant variance for the random effect ($\widehat{\sigma^2} = 0.985$). Second, we use equations (9) and (12). The credibility is not a deterministic function of $\widehat{\lambda}_{i1}$, due to the second component in equation (12). We averaged it with a Gaussian kernel. The optimal bandwidth is small ($h = 0.001$) because the variations around the average are very low. As expected, the credibility is almost constant with the power link between the variance of the random effect and the frequency premium because the estimated elasticity is close to -1.

6 Conclusions

The relative variations after each year for coefficients of real-world bonus-malus scales do not depend on the frequency premium (however they depend on the location in the scale). Relaxing the equidistribution assumption on the random effects may allow actuarial models to get closer to real-world rating structures concerning the bonus, if the estimated link between the variance of the random effect and frequency risk is decreasing. As an actuarial malus is close to the variance of the random effect if risk exposure is low, the malus also decreases with the frequency premium if the aforementioned link is decreasing. Hence a real-world bonus-malus scale is close to a usual actuarial model on the malus side, whereas on the bonus side it can be closer to the models developed in this paper.

A Appendix: Consistency of the estimation under link violation

This appendix gives the conditions which provide a consistent estimation of the conditional expectation $E(Y | X)$ in a misspecification context on the link between the index and $E(Y)$. The reference is Li and Duan (1989).

Suppose that we have a sampling model on the pair (x, y) , with the nonparametric link $E(Y | X = x) = f(x'\beta)$. Notations are those of Section 2. In the aforementioned reference, the distribution of Y given $X = x$ is supposed to be that of $g(x'\beta, \varepsilon)$, where ε is a given random variable and g a given function. This implicitly assumes that the distribution of Y is determined by the expectation $E(Y)$, which is the case for binary or Poisson distributions.

Besides, a parametric model with a scalar parameter m and a likelihood L_m is estimated on the data (observations are $(x_i, y_i)_{i=1, \dots, n}$). The scalar parameter is linked to the index $x'b$ by a given function f_0 . We write

$$m_i = f_0(c + x'_i b); \quad -\log L_{m_i}(y_i) = h(c + x'_i b, y_i).$$

For instance, in a Poisson model the parameter is the expectation and the exponential link is usually retained for f_0 . In that case we have: $h(s, y) = \exp(s) - ys + \log(y!)$.

Let $\hat{c}, \hat{b} = \arg \min_{c, b} \sum_i h(c + x'_i b, y_i)$ be the maximum likelihood estimators in the parametric model. If data are generated by the sampling model given in the first place, this estimator converges towards a limit c_0, b_0 usually called a pseudo-true value (Gouriéroux et al., 1984). We have that

$$b_0 \in \mathbb{R}\beta \text{ (i.e. } b_0 = \lambda\beta, \lambda \in \mathbb{R}) \tag{13}$$

under the following assumptions.

1. The maps $h(\bullet, y) : s \rightarrow h(s, y)$ are convex for every value of y . This assumption implies that the map $c, b \rightarrow E[h(c + X'b, Y)] = R(c, b)$ is convex.
2. The minimum of the map R defined in Assumption 1 is reached for only one pair c_0, b_0 .
3. We have the following property

$$\forall b \in \mathbb{R}^k, \exists d, \lambda \in \mathbb{R}, E(X'b | X'\beta) = d + (\lambda X'\beta). \tag{14}$$

Let us comment the result and the assumptions. Since we have the property $\hat{b} \xrightarrow{a.e.} b_0 = \lambda\beta$, the estimation of $E(Y | X)$ obtained from (3) and (4) will be consistent. Indeed, it is the line $\mathbb{R}\beta$ which must be identified in an index model, as discussed in Section 2. Hence the multiplicative constant λ does not matter in the estimation.

The first assumption relates to the concavity of the log-likelihood and is usually fulfilled. The consistency property of the maximum likelihood estimation and assumption 2 lead to

$$\hat{c}, \hat{b} = \arg \min_{c,b} \frac{1}{n} \sum_{i=1}^n h(c + x_i' b, y_i) \xrightarrow{a.e.} c_0, b_0 = \arg \min_{c,b} R(c, b) = E \left[h(c + X' b, Y) \right].$$

From the strong law of large numbers, it is easily seen why assumption 2 is necessary.

The consistency result given in (13) is obtained from (14) with assumption 2 and Jensen's inequality applied on the functions $h(\bullet, y)$. Assumption 3 is generally not fulfilled for variables X with discrete values, as it is the case in our empirical analysis. Equation (14) is fulfilled if X follows a non degenerate Gaussian distribution. Indeed, it is well known that in that case the conditional expectation defined in (14) is obtained from the affine probabilistic regression of $X'b$ with respect to $X'\beta$. Property (14) is more generally fulfilled for elliptically symmetric distributions.

References

- [1] Bolancé, C., Guillén, M., Pinquet, J., 2003. Time-varying credibility for frequency risk models: estimation and tests for autoregressive specifications on the random effects. *Insurance: Mathematics and Economics* 33, 273-282.
- [2] Brouhns, N., Guillén, M., Denuit, M., Pinquet, J., 2003. Bonus-malus scales in segmented tariffs with stochastic migration between segments. *The Journal of Risk and Insurance* 70, 577-599.

- [3] Bühlmann, H., 1967. Experience rating and credibility. ASTIN Bulletin 4, 199-207.
- [4] Dionne, G., Vanasse, C., 1989. A generalization of automobile insurance rating models: the negative binomial distribution with a regression component. ASTIN Bulletin 19, 199-212.
- [5] Dufresne, F., Gerber H., Shiu E., 1991. Risk theory with the gamma process. ASTIN Bulletin 21, 177-192.
- [6] Gouriéroux, C., Monfort, A., Trognon, A., 1984. Pseudo likelihood methods: Theory. Econometrica 52, 681-700.
- [7] Härdle, W., 1990. Applied nonparametric regression. Econometric Society Monographs, Cambridge University Press.
- [8] Härdle, W., Hall, P., Ichimura, H., 1993. Optimal smoothing in single index models. The Annals of Statistics 21, 157-178.
- [9] Härdle, W., Spokoiny, V., Sperlich, S., 1997. Semiparametric single index versus fixed link function modelling. The Annals of Statistics 25, 212-243.
- [10] Harvey, A.C., 1976. Estimating regression models with multiplicative heteroscedasticity. Econometrica 44, 461-465.
- [11] Li, K.C., Duan, N., 1989. Regression analysis under link violation. The Annals of Statistics 21, 157-178.
- [12] Qian, W., 2000. An application of nonparametric regression estimation in credibility theory. Insurance: Mathematics and Economics 27, 169-176.
- [13] Winkelmann, R., Zimmermann, K.F., 1991. A new approach for modelling economic count data. Economic Letters 37, 139-143.